

```
In [1]: import pandas as pd
import numpy as np
import scipy
from scipy.spatial import distance
import math
from datetime import datetime
from dateutil import parser
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline
import os
from scipy import stats
sns.set_style("darkgrid")
import matplotlib.image as mpimg
import matplotlib.ticker as ticker
from IPython.display import IFrame
import folium
from folium import plugins
from folium.plugins import MarkerCluster, FastMarkerCluster, HeatMapWith
Time,HeatMap
#!pip install git+https://github.com/python-visualization/folium
```

```
In [2]: cdata_19=pd.read_csv('Crimes_-_2019.csv')
cdata_18=pd.read_csv('Crimes_-_2018.csv')
pdata=pd.read_csv('Police_Stations.csv')
```

```
In [3]: cdata_18.head()
```

Out[3]:

|   | ID       | Case<br>Number | Date                         | Block                             | IUCR | Primary<br>Type                  | Description                         | Location<br>Description |
|---|----------|----------------|------------------------------|-----------------------------------|------|----------------------------------|-------------------------------------|-------------------------|
| 0 | 11903290 | JC508826       | 01/01/2018<br>08:00:00<br>AM | 085XX S<br>ESCANABA<br>AVE        | 1752 | OFFENSE<br>INVOLVING<br>CHILDREN | AGG CRIM<br>SEX ABUSE<br>FAM MEMBER | RESIDENCE               |
| 1 | 11902580 | JC524113       | 09/01/2018<br>06:00:00<br>AM | 094XX S<br>MICHIGAN<br>AVE        | 2825 | OTHER<br>OFFENSE                 | HARASSMENT<br>BY<br>TELEPHONE       | RESIDENCE               |
| 2 | 11516628 | JB528220       | 11/24/2018<br>08:45:00<br>PM | 095XX S<br>STONY<br>ISLAND<br>AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT        | AGGRAVATED:<br>HANDGUN              | ALLEY                   |
| 3 | 11491979 | JB495848       | 08/27/2018<br>12:00:00<br>PM | 092XX S<br>UNION AVE              | 1120 | DECEPTIVE<br>PRACTICE            | FORGERY                             | RESIDENCE               |
| 4 | 11895996 | JC516056       | 04/01/2018<br>12:00:00<br>AM | 003XX W<br>CHESTNUT<br>ST         | 0810 | THEFT                            | OVER \$500                          | APARTMENT               |

5 rows × 22 columns

```
In [4]: cdata_19.head()
```

```
Out[4]:
```

|   | ID       | Case Number | Date              | Block                        | IUCR | Primary Type          | Description                   | Location Description          | Arr |
|---|----------|-------------|-------------------|------------------------------|------|-----------------------|-------------------------------|-------------------------------|-----|
| 0 | 11889339 | JC507601    | 11-12-19<br>12:15 | 025XX W<br>62ND ST           | 486  | BATTERY               | DOMESTIC<br>BATTERY<br>SIMPLE | RESIDENCE                     | Fa  |
| 1 | 11893558 | JC513193    | 11-09-19<br>18:00 | 011XX E<br>83RD ST           | 820  | THEFT                 | \$500 AND<br>UNDER            | VEHICLE<br>NON-<br>COMMERCIAL | Fa  |
| 2 | 11889659 | JC508257    | 11-12-19<br>18:45 | 003XX E<br>BENTON PL         | 2820 | OTHER<br>OFFENSE      | TELEPHONE<br>THREAT           | SMALL<br>RETAIL<br>STORE      | Fa  |
| 3 | 11891583 | JC510501    | 11-04-19<br>20:09 | 020XX S<br>MORGAN<br>ST      | 1122 | DECEPTIVE<br>PRACTICE | COUNTERFEIT<br>CHECK          | APARTMENT                     | Fa  |
| 4 | 11890017 | JC508355    | 11-12-19<br>21:14 | 040XX S<br>CALIFORNIA<br>AVE | 560  | ASSAULT               | SIMPLE                        | OTHER                         | Fa  |

5 rows × 22 columns

```
In [5]: cdata_18.shape
```

```
Out[5]: (267639, 22)
```

```
In [6]: cdata_19.shape
```

```
Out[6]: (224412, 22)
```

```
In [7]: #merge=cdata.merge(pdata,how='inner',left_on=['District'],right_on=['District'])
#Task 2
#merge=pd.merge(cdata,pdata,how='inner', left_on=['District'],right_on=
['District'])
cdata=pd.concat([cdata_18,cdata_19]).reset_index()
```

```
In [8]: cdata.shape
```

```
Out[8]: (492051, 23)
```

```
In [9]: cdata.head()
```

```
Out[9]:
```

|   | index | ID       | Case Number | Date                      | Block                       | IUCR | Primary Type                     | Description                         | Loc Descri |
|---|-------|----------|-------------|---------------------------|-----------------------------|------|----------------------------------|-------------------------------------|------------|
| 0 | 0     | 11903290 | JC508826    | 01/01/2018<br>08:00:00 AM | 085XX S<br>ESCANABA AVE     | 1752 | OFFENSE<br>INVOLVING<br>CHILDREN | AGG CRIM<br>SEX ABUSE<br>FAM MEMBER | RESIDE     |
| 1 | 1     | 11902580 | JC524113    | 09/01/2018<br>06:00:00 AM | 094XX S<br>MICHIGAN AVE     | 2825 | OTHER<br>OFFENSE                 | HARASSMENT<br>BY<br>TELEPHONE       | RESIDE     |
| 2 | 2     | 11516628 | JB528220    | 11/24/2018<br>08:45:00 PM | 095XX S<br>STONY ISLAND AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT        | AGGRAVATED:<br>HANDGUN              | A          |
| 3 | 3     | 11491979 | JB495848    | 08/27/2018<br>12:00:00 PM | 092XX S<br>UNION AVE        | 1120 | DECEPTIVE<br>PRACTICE            | FORGERY                             | RESIDE     |
| 4 | 4     | 11895996 | JC516056    | 04/01/2018<br>12:00:00 AM | 003XX W<br>CHESTNUT ST      | 0810 | THEFT                            | OVER \$500                          | APARTM     |

5 rows × 23 columns

```
In [10]: pdata.head()
```

```
Out[10]:
```

|   | DISTRICT     | DISTRICT NAME  | ADDRESS                 | CITY    | STATE | ZIP   |   |
|---|--------------|----------------|-------------------------|---------|-------|-------|---|
| 0 | Headquarters | Headquarters   | 3510 S<br>Michigan Ave  | Chicago | IL    | 60653 | <a href="http://home.chicagopolice.org/correctional">http://home.chicagopolice.org/correctional</a> |
| 1 | 1            | Central        | 1718 S<br>State St      | Chicago | IL    | 60616 | <a href="http://home.chicagopolice.org/correctional">http://home.chicagopolice.org/correctional</a> |
| 2 | 6            | Gresham        | 7808 S<br>Halsted St    | Chicago | IL    | 60620 | <a href="http://home.chicagopolice.org/correctional">http://home.chicagopolice.org/correctional</a> |
| 3 | 11           | Harrison       | 3151 W<br>Harrison St   | Chicago | IL    | 60612 | <a href="http://home.chicagopolice.org/correctional">http://home.chicagopolice.org/correctional</a> |
| 4 | 16           | Jefferson Park | 5151 N<br>Milwaukee Ave | Chicago | IL    | 60630 | <a href="http://home.chicagopolice.org/correctional">http://home.chicagopolice.org/correctional</a> |

```
In [11]: pdata=pdata.drop(columns=['ADDRESS', 'CITY', 'STATE', 'ZIP', 'WEBSITE', 'PHONE', 'FAX', 'TTY', 'LOCATION'])
```

```
In [12]: cdata.nunique()
```

```
Out[12]: index                267639
ID                492051
Case Number       492007
Date              233022
Block             30852
IUCR              423
Primary Type      32
Description       315
Location Description 137
Arrest            2
Domestic          2
Beat             274
District          23
Ward              50
Community Area    78
FBI Code          32
X Coordinate      57981
Y Coordinate      88263
Year              2
Updated On        1439
Latitude          232807
Longitude         232764
Location          189633
dtype: int64
```

```
In [13]: pdata.head()
```

```
Out[13]:
```

|   | DISTRICT     | DISTRICT NAME  | X COORDINATE | Y COORDINATE | LATITUDE  | LONGITUDE  |
|---|--------------|----------------|--------------|--------------|-----------|------------|
| 0 | Headquarters | Headquarters   | 1177731.401  | 1881697.404  | 41.830702 | -87.623395 |
| 1 | 1            | Central        | 1176569.052  | 1891771.704  | 41.858373 | -87.627356 |
| 2 | 6            | Gresham        | 1172283.013  | 1853022.646  | 41.752137 | -87.644229 |
| 3 | 11           | Harrison       | 1155244.069  | 1897148.755  | 41.873582 | -87.705488 |
| 4 | 16           | Jefferson Park | 1138480.758  | 1933660.473  | 41.974094 | -87.766149 |

```
In [14]: pdata.rename(columns={'DISTRICT': 'District', 'DISTRICT NAME': 'p_dname', 'X
COORDINATE': 'p_x', 'Y COORDINATE': 'p_y', 'LATITUDE': 'p_latitude', 'LONGITUD
E': 'p_longitude'}, inplace=True)
```

```
In [15]: pdata.head()
pdata=pdata[pdata.District!='Headquarters']
```

```
In [16]: #No of arrests and non arrests
cdata['Arrest'].value_counts()
```

```
Out[16]: False    391229
True      100822
Name: Arrest, dtype: int64
```

```
In [17]: #Count of each type of crime  
cdata['Primary Type'].value_counts()
```

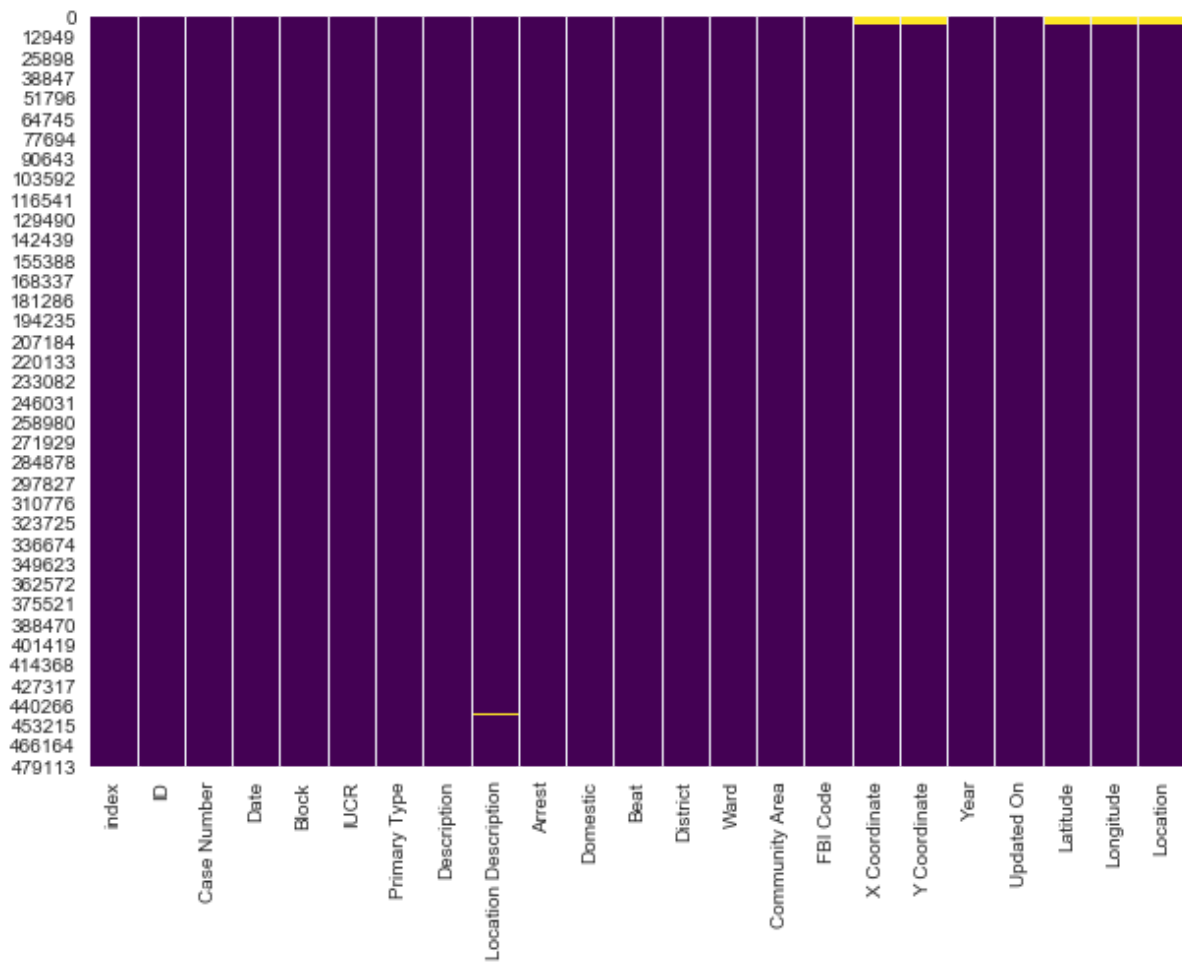
```
Out[17]: THEFT 118793  
BATTERY 93202  
CRIMINAL DAMAGE 51276  
ASSAULT 38551  
DECEPTIVE PRACTICE 34293  
OTHER OFFENSE 31796  
NARCOTICS 25611  
BURGLARY 20010  
MOTOR VEHICLE THEFT 17744  
ROBBERY 16530  
CRIMINAL TRESPASS 12902  
WEAPONS VIOLATION 10979  
OFFENSE INVOLVING CHILDREN 4273  
CRIM SEXUAL ASSAULT 3015  
PUBLIC PEACE VIOLATION 2754  
INTERFERENCE WITH PUBLIC OFFICER 2677  
SEX OFFENSE 2199  
PROSTITUTION 1294  
HOMICIDE 1039  
ARSON 692  
LIQUOR LAW VIOLATION 479  
STALKING 399  
CONCEALED CARRY LICENSE VIOLATION 347  
GAMBLING 338  
INTIMIDATION 318  
KIDNAPPING 313  
OBSCENITY 131  
NON-CRIMINAL 39  
HUMAN TRAFFICKING 24  
PUBLIC INDECENCY 23  
OTHER NARCOTIC VIOLATION 7  
NON-CRIMINAL (SUBJECT SPECIFIED) 3  
Name: Primary Type, dtype: int64
```

```
In [18]: #Find missing values  
cdata.isnull().sum()
```

```
Out[18]: index                                0  
ID                                              0  
Case Number                                   0  
Date                                           0  
Block                                          0  
IUCR                                           0  
Primary Type                                 0  
Description                                   0  
Location Description                         1748  
Arrest                                        0  
Domestic                                     0  
Beat                                          0  
District                                     2  
Ward                                          15  
Community Area                              5  
FBI Code                                    0  
X Coordinate                               5241  
Y Coordinate                               5241  
Year                                         0  
Updated On                                 0  
Latitude                                   5241  
Longitude                                   5241  
Location                                   5241  
dtype: int64
```

```
In [19]: plt.figure(figsize=(10,7))
sns.heatmap(cdata.isnull(), cbar = False, cmap = 'viridis')
```

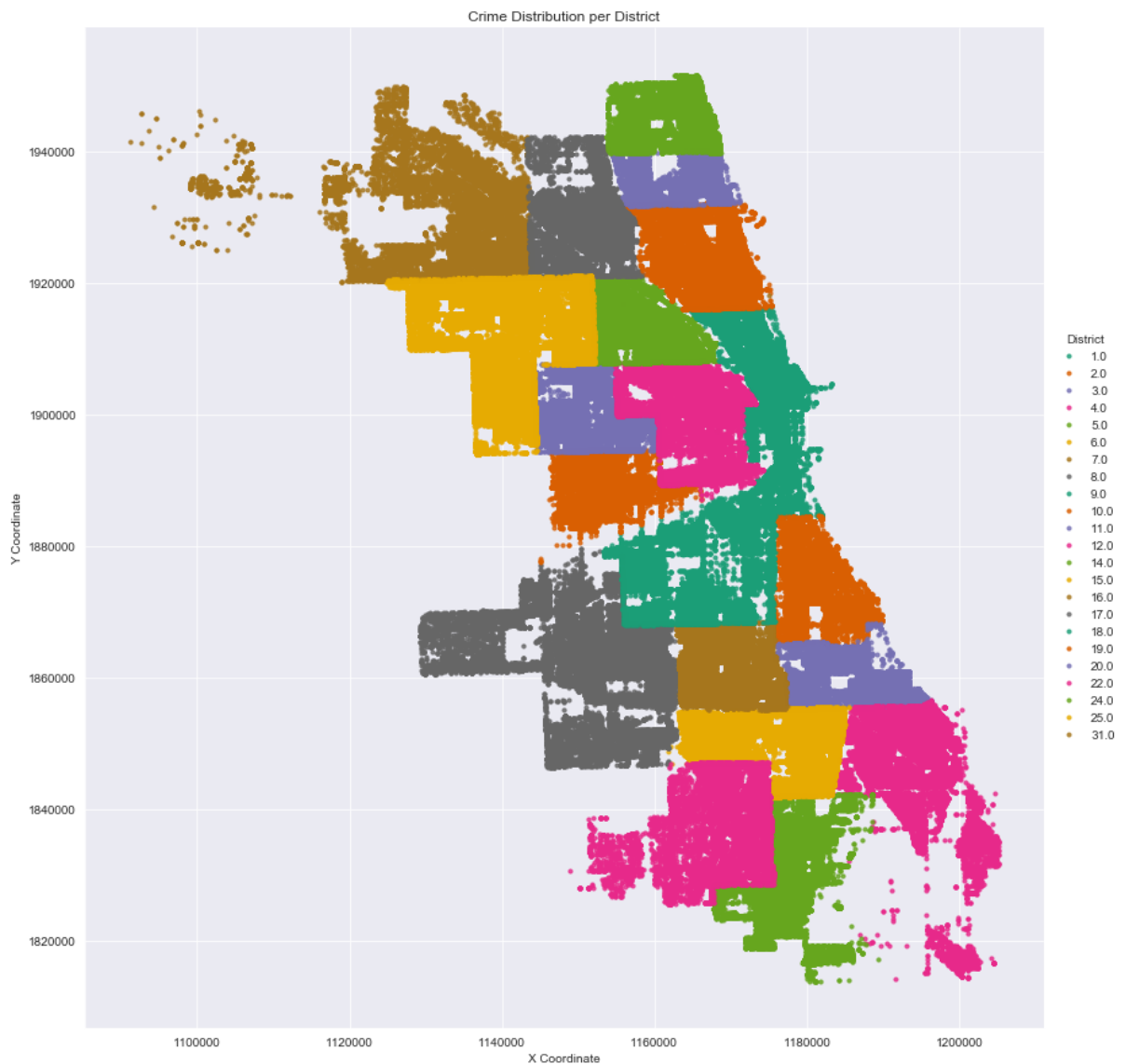
```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1c16c5d0>
```



```
In [20]: cdata = cdata.loc[(cdata['X Coordinate']!=0)]
sns.lmplot('X Coordinate',
           'Y Coordinate',
           data=cdata[:],
           fit_reg=False,
           hue="District", #District
           palette='Dark2',
           height=12,
           ci=2,
           scatter_kws={"marker": "D",
                        "s": 10})

ax = plt.gca()
ax.set_title("Crime Distribution per District")
```

```
Out[20]: Text(0.5, 1, 'Crime Distribution per District')
```





```
In [21]: #Dropping rows with null values
cdata=cdata[pd.notnull(cdata['Location Description'])]
cdata=cdata[pd.notnull(cdata['District'])]
cdata=cdata[pd.notnull(cdata['Ward'])]
cdata=cdata[pd.notnull(cdata['X Coordinate'])]
cdata=cdata[pd.notnull(cdata['Community Area'])]
```

```
In [22]: cdata.isnull().sum()
```

```
Out[22]: index          0
ID          0
Case Number  0
Date         0
Block        0
IUCR         0
Primary Type  0
Description   0
Location Description  0
Arrest        0
Domestic      0
Beat          0
District      0
Ward          0
Community Area  0
FBI Code      0
X Coordinate   0
Y Coordinate   0
Year           0
Updated On     0
Latitude       0
Longitude      0
Location       0
dtype: int64
```

```
In [23]: cdata.iloc[:5,[20,21]]
```

```
Out[23]:
```

|   | Latitude  | Longitude  |
|---|-----------|------------|
| 2 | 41.721844 | -87.585072 |
| 3 | 41.726124 | -87.640913 |
| 5 | 41.770428 | -87.628300 |
| 6 | 41.992509 | -87.660344 |
| 9 | 41.765884 | -87.693061 |

```
In [24]: chicago = [41.85, -87.68]
m = folium.Map(location=chicago, zoom_start=9.5, control_scale = False)
plugins.Fullscreen(
    position='topright',
    title='Expand me',
    title_cancel='Exit me',
    force_separate_button=True).add_to(m)
folium.Choropleth(
    geo_data='chicago_police_districts.geojson',
    name='Choropleth',
    data=cdata,
    #data=cdata[cdata['Primary Type']=='ASSAULT'].dropna(),
    columns=['District', 'Arrest'],
    key_on='feature.properties.dist_num',
    fill_color='YlOrRd',
    fill_opacity=0.4,
    line_opacity=0.2,
    legend_name='Arrests',
    highlight=True
).add_to(m)
m.add_child(plugins.HeatMap(cdata.iloc[:, [20, 21]],
                           name='Assault',
                           radius=5,
                           max_zoom=4,
                           blur=10
                           ))
folium.TileLayer('openstreetmap').add_to(m)
folium.TileLayer('cartodbpositron').add_to(m)
folium.LayerControl().add_to(m)
m.save("map3.html")
IFrame('map3.html', width=990, height=700)
```

```
<string>:6: RuntimeWarning:
```

```
Converting input from bool to <class 'numpy.uint8'> for compatibility.
```

```
Out[24]:
```

```
In [25]: cdata['District'] = cdata.District.astype(int)
         pdata['District'] = pdata.District.astype(int)
```

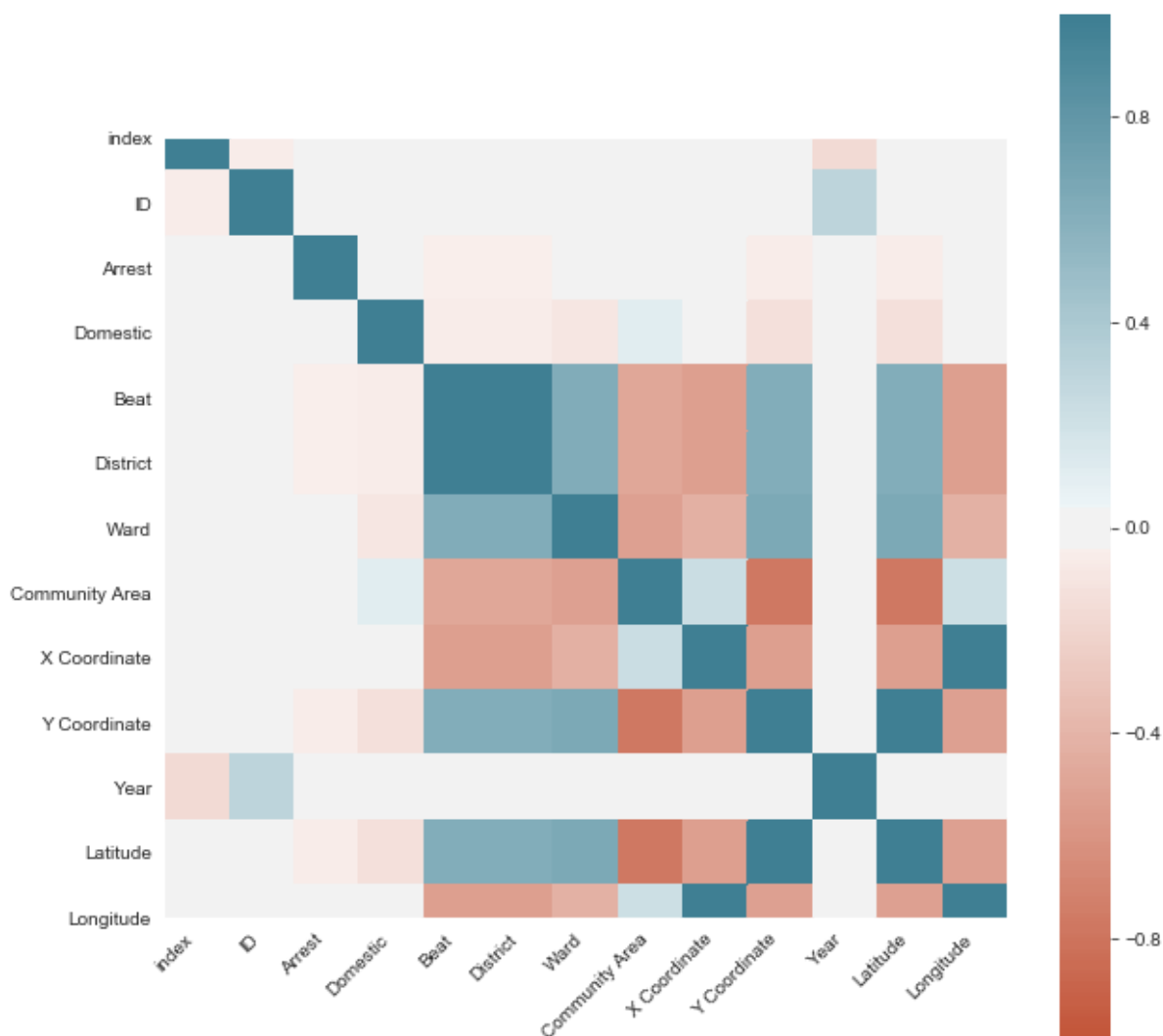
```
In [26]: pdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 22 entries, 1 to 22
Data columns (total 6 columns):
District      22 non-null int64
p_dname       22 non-null object
p_x           22 non-null float64
p_y           22 non-null float64
p_latitude    22 non-null float64
p_longitude   22 non-null float64
dtypes: float64(4), int64(1), object(1)
memory usage: 1.2+ KB
```

```
In [27]: cdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 485419 entries, 2 to 492050
Data columns (total 23 columns):
index          485419 non-null int64
ID             485419 non-null int64
Case Number    485419 non-null object
Date           485419 non-null object
Block          485419 non-null object
IUCR           485419 non-null object
Primary Type   485419 non-null object
Description     485419 non-null object
Location Description 485419 non-null object
Arrest         485419 non-null bool
Domestic       485419 non-null bool
Beat           485419 non-null int64
District       485419 non-null int64
Ward           485419 non-null float64
Community Area 485419 non-null float64
FBI Code       485419 non-null object
X Coordinate    485419 non-null float64
Y Coordinate    485419 non-null float64
Year           485419 non-null int64
Updated On     485419 non-null object
Latitude        485419 non-null float64
Longitude       485419 non-null float64
Location        485419 non-null object
dtypes: bool(2), float64(6), int64(5), object(10)
memory usage: 82.4+ MB
```

```
In [28]: corr = cdata.corr()
fig, ax = plt.subplots(figsize=(10,10))
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);
```



```
In [29]: cdata.shape
```

```
Out[29]: (485419, 23)
```

```
In [30]: pdata.shape
```

```
Out[30]: (22, 6)
```

```
In [31]: pdata['District'].unique()  
#
```

```
Out[31]: array([ 1,  6, 11, 16, 24,  2,  7, 25, 10, 15,  3, 19, 14,  8,  4, 20,  
               18,  
               12,  9, 22,  5, 17])
```

```
In [32]: cdata['District'].unique()
```

```
Out[32]: array([ 4, 22,  7, 24,  8,  2, 18, 15,  9, 12,  1, 25,  3, 11, 14, 16,  
               10,  
               6, 19, 17, 31,  5, 20])
```

```
In [33]: cdata[cdata["District"]==31].count()
```

```
Out[33]: index                10  
ID                10  
Case Number       10  
Date              10  
Block             10  
IUCR              10  
Primary Type      10  
Description        10  
Location Description 10  
Arrest            10  
Domestic          10  
Beat             10  
District          10  
Ward              10  
Community Area    10  
FBI Code          10  
X Coordinate      10  
Y Coordinate      10  
Year              10  
Updated On        10  
Latitude          10  
Longitude         10  
Location          10  
dtype: int64
```

```
In [34]: cdata.loc[cdata['District']==31]
```

Out[34]:

|        | index  | ID       | Case Number | Date                      | Block                       | IUCR | Primary Type               | Descri                 |
|--------|--------|----------|-------------|---------------------------|-----------------------------|------|----------------------------|------------------------|
| 177    | 177    | 11513541 | JB524308    | 08/01/2018<br>12:01:00 AM | 037XX S<br>HERMITAGE AVE    | 1754 | OFFENSE INVOLVING CHILDREN | AGG SEX ASSI CHILD FAM |
| 281    | 281    | 11498874 | JB504419    | 10/27/2018<br>08:10:00 PM | 008XX E<br>89TH ST          | 1544 | SEX OFFENSE                | SE EXPLOITATION C      |
| 306    | 306    | 11315204 | JB256739    | 01/01/2018<br>12:01:00 AM | 013XX N<br>BELL AVE         | 1544 | SEX OFFENSE                | SE EXPLOITATION C      |
| 307    | 307    | 11706302 | JC285535    | 03/07/2018<br>12:00:00 AM | 072XX S<br>INGLESIDE AVE    | 1565 | SEX OFFENSE                | INDEI SOLICITATION/C   |
| 108022 | 108022 | 11336318 | JB291739    | 06/03/2018<br>09:00:00 PM | 026XX N<br>LAKE SHORE DR NB | 041A | BATTERY                    | AGGRAV/ HANE           |
| 114728 | 114728 | 11345502 | JB303991    | 06/08/2018<br>06:00:00 PM | 052XX N<br>WESLEY TER       | 0910 | MOTOR VEHICLE THEFT        | AUTOMC                 |
| 176090 | 176090 | 11428268 | JB410717    | 08/23/2018<br>09:00:00 AM | 115XX W<br>TOUHY AVE        | 1130 | DECEPTIVE PRACTICE         | FRAU CONFIDENCE C      |
| 205229 | 205229 | 11467984 | JB464102    | 06/06/2018<br>11:24:00 AM | 003XX S<br>WHIPPLE ST       | 0266 | CRIM SEXUAL ASSAULT        | PREDAS                 |
| 257204 | 257204 | 11539710 | JB558892    | 12/16/2018<br>07:00:00 PM | 026XX S<br>DRAKE AVE        | 1754 | OFFENSE INVOLVING CHILDREN | AGG SEX ASSI CHILD FAM |
| 304332 | 36693  | 11842446 | JC450320    | 09-10-19<br>9:00          | 068XX N<br>OLMSTED AVE      | 890  | THEFT                      | FROM BUIL              |

10 rows × 23 columns

```
In [35]: cdata=cdata[cdata.District!=31]
```

```
In [36]: cdata.loc[cdata['District']==31]
```

Out[36]:

| index | ID | Case Number | Date | Block | IUCR | Primary Type | Description | Location Description | Arrest | ... | Ward | Cc |
|-------|----|-------------|------|-------|------|--------------|-------------|----------------------|--------|-----|------|----|
|-------|----|-------------|------|-------|------|--------------|-------------|----------------------|--------|-----|------|----|

0 rows × 23 columns

```
In [37]: cdata.iloc[0:5,8:14]
```

```
Out[37]:
```

|   | Location Description         | Arrest | Domestic | Beat | District | Ward |
|---|------------------------------|--------|----------|------|----------|------|
| 2 | ALLEY                        | True   | False    | 431  | 4        | 7.0  |
| 3 | RESIDENCE                    | True   | False    | 2223 | 22       | 21.0 |
| 5 | VEHICLE NON-COMMERCIAL       | True   | False    | 722  | 7        | 6.0  |
| 6 | COMMERCIAL / BUSINESS OFFICE | True   | False    | 2433 | 24       | 48.0 |
| 9 | APARTMENT                    | False  | True     | 831  | 8        | 17.0 |

```
In [38]: pdata.reset_index(drop=True,inplace=True)
```

```
In [39]: pdata.head()
```

```
Out[39]:
```

|   | District | p_dname        | p_x         | p_y         | p_latitude | p_longitude |
|---|----------|----------------|-------------|-------------|------------|-------------|
| 0 | 1        | Central        | 1176569.052 | 1891771.704 | 41.858373  | -87.627356  |
| 1 | 6        | Gresham        | 1172283.013 | 1853022.646 | 41.752137  | -87.644229  |
| 2 | 11       | Harrison       | 1155244.069 | 1897148.755 | 41.873582  | -87.705488  |
| 3 | 16       | Jefferson Park | 1138480.758 | 1933660.473 | 41.974094  | -87.766149  |
| 4 | 24       | Rogers Park    | 1164193.588 | 1943199.401 | 41.999763  | -87.671324  |

```
In [40]: #merge=cdata.merge(pdata,how='inner',left_on=['District'],right_on=['District'])
#Task 2
#merge=pd.merge(cdata,pdata,how='inner', left_on=['District'],right_on=['District'])
merge=pd.concat([cdata.set_index('District'),pdata.set_index('District')], axis=1, join='inner').reset_index()
```



```
In [41]: merge.head()
```

```
Out[41]:
```

|   | District | index | ID       | Case<br>Number | Date                         | Block                          | IUCR | Primary<br>Type           | Description                        |
|---|----------|-------|----------|----------------|------------------------------|--------------------------------|------|---------------------------|------------------------------------|
| 0 | 4        | 2     | 11516628 | JB528220       | 11/24/2018<br>08:45:00<br>PM | 095XX S<br>STONY<br>ISLAND AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT | AGGRAVATED<br>HANDGUN              |
| 1 | 22       | 3     | 11491979 | JB495848       | 08/27/2018<br>12:00:00<br>PM | 092XX S<br>UNION AVE           | 1120 | DECEPTIVE<br>PRACTICE     | FORGERY                            |
| 2 | 7        | 5     | 11499634 | JB505888       | 09/27/2018<br>10:30:00<br>AM | 068XX S<br>PERRY AVE           | 1020 | ARSON                     | BY FIRE                            |
| 3 | 24       | 6     | 11474586 | JB472665       | 10/12/2018<br>12:04:00<br>PM | 061XX N<br>BROADWAY            | 0890 | THEFT                     | FROM<br>BUILDING                   |
| 4 | 8        | 9     | 11298701 | JB241744       | 04/27/2018<br>02:45:00<br>PM | 070XX S<br>CALIFORNIA<br>AVE   | 1562 | SEX<br>OFFENSE            | AGG<br>CRIMINAL<br>SEXUAL<br>ABUSE |

5 rows × 28 columns

```
In [42]: merge.shape
```

```
Out[42]: (485409, 28)
```

```
In [43]: merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 485409 entries, 0 to 485408
Data columns (total 28 columns):
District                485409 non-null int64
index                   485409 non-null int64
ID                      485409 non-null int64
Case Number             485409 non-null object
Date                   485409 non-null object
Block                  485409 non-null object
IUCR                   485409 non-null object
Primary Type           485409 non-null object
Description             485409 non-null object
Location Description    485409 non-null object
Arrest                 485409 non-null bool
Domestic               485409 non-null bool
Beat                  485409 non-null int64
Ward                   485409 non-null float64
Community Area         485409 non-null float64
FBI Code               485409 non-null object
X Coordinate           485409 non-null float64
Y Coordinate           485409 non-null float64
Year                   485409 non-null int64
Updated On            485409 non-null object
Latitude              485409 non-null float64
Longitude             485409 non-null float64
Location              485409 non-null object
p_dname               485409 non-null object
p_x                   485409 non-null float64
p_y                   485409 non-null float64
p_latitude            485409 non-null float64
p_longitude           485409 non-null float64
dtypes: bool(2), float64(10), int64(5), object(11)
memory usage: 97.2+ MB
```

```
In [44]: #Dropping irrelevant columns
merge=merge.drop(columns=['ID', 'Updated On', 'Year', 'Location'])
```

```
In [45]: merge.shape
merge.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 485409 entries, 0 to 485408
Data columns (total 24 columns):
District                485409 non-null int64
index                  485409 non-null int64
Case Number             485409 non-null object
Date                   485409 non-null object
Block                  485409 non-null object
IUCR                   485409 non-null object
Primary Type           485409 non-null object
Description             485409 non-null object
Location Description    485409 non-null object
Arrest                 485409 non-null bool
Domestic               485409 non-null bool
Beat                   485409 non-null int64
Ward                   485409 non-null float64
Community Area         485409 non-null float64
FBI Code               485409 non-null object
X Coordinate           485409 non-null float64
Y Coordinate           485409 non-null float64
Latitude               485409 non-null float64
Longitude              485409 non-null float64
p_dname                485409 non-null object
p_x                    485409 non-null float64
p_y                    485409 non-null float64
p_latitude             485409 non-null float64
p_longitude            485409 non-null float64
dtypes: bool(2), float64(10), int64(3), object(9)
memory usage: 82.4+ MB
```

```
In [46]: # Changing Arrest's to 0-False and 1-True
change_values = {True :1,False :0}
merge['Arrest'] = merge['Arrest'].map(change_values)
merge['Domestic']=merge['Domestic'].map(change_values)
```

```
In [47]: merge.head()
```

```
Out[47]:
```

|   | District | index | Case Number | Date                         | Block                          | IUCR | Primary Type              | Description                        | Loc Descri             |
|---|----------|-------|-------------|------------------------------|--------------------------------|------|---------------------------|------------------------------------|------------------------|
| 0 | 4        | 2     | JB528220    | 11/24/2018<br>08:45:00<br>PM | 095XX S<br>STONY<br>ISLAND AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT | AGGRAVATED:<br>HANDGUN             | A                      |
| 1 | 22       | 3     | JB495848    | 08/27/2018<br>12:00:00<br>PM | 092XX S<br>UNION AVE           | 1120 | DECEPTIVE<br>PRACTICE     | FORGERY                            | RESIDI                 |
| 2 | 7        | 5     | JB505888    | 09/27/2018<br>10:30:00<br>AM | 068XX S<br>PERRY AVE           | 1020 | ARSON                     | BY FIRE                            | VEH<br>COMMEF          |
| 3 | 24       | 6     | JB472665    | 10/12/2018<br>12:04:00<br>PM | 061XX N<br>BROADWAY            | 0890 | THEFT                     | FROM<br>BUILDING                   | COMMEF<br>/ BUSI<br>OI |
| 4 | 8        | 9     | JB241744    | 04/27/2018<br>02:45:00<br>PM | 070XX S<br>CALIFORNIA<br>AVE   | 1562 | SEX<br>OFFENSE            | AGG<br>CRIMINAL<br>SEXUAL<br>ABUSE | APARTI                 |

5 rows × 24 columns

```
In [48]: merge['Distance']=np.sqrt((merge['X Coordinate']-merge['p_x'])**2 + (mer  
ge['Y Coordinate']-merge['p_y'])**2)
```

```
In [49]: merge.head()
```

```
Out[49]:
```

|   | District | index | Case Number | Date                         | Block                          | IUCR | Primary Type              | Description                        | Loc Descri             |
|---|----------|-------|-------------|------------------------------|--------------------------------|------|---------------------------|------------------------------------|------------------------|
| 0 | 4        | 2     | JB528220    | 11/24/2018<br>08:45:00<br>PM | 095XX S<br>STONY<br>ISLAND AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT | AGGRAVATED:<br>HANDGUN             | A                      |
| 1 | 22       | 3     | JB495848    | 08/27/2018<br>12:00:00<br>PM | 092XX S<br>UNION AVE           | 1120 | DECEPTIVE<br>PRACTICE     | FORGERY                            | RESIDI                 |
| 2 | 7        | 5     | JB505888    | 09/27/2018<br>10:30:00<br>AM | 068XX S<br>PERRY AVE           | 1020 | ARSON                     | BY FIRE                            | VEH<br>COMMEF          |
| 3 | 24       | 6     | JB472665    | 10/12/2018<br>12:04:00<br>PM | 061XX N<br>BROADWAY            | 0890 | THEFT                     | FROM<br>BUILDING                   | COMMEF<br>/ BUSI<br>OI |
| 4 | 8        | 9     | JB241744    | 04/27/2018<br>02:45:00<br>PM | 070XX S<br>CALIFORNIA<br>AVE   | 1562 | SEX<br>OFFENSE            | AGG<br>CRIMINAL<br>SEXUAL<br>ABUSE | APARTI                 |

5 rows × 25 columns

```
In [50]: merge[ 'Date' ][0:5]
```

```
Out[50]: 0    11/24/2018 08:45:00 PM
          1    08/27/2018 12:00:00 PM
          2    09/27/2018 10:30:00 AM
          3    10/12/2018 12:04:00 PM
          4    04/27/2018 02:45:00 PM
          Name: Date, dtype: object
```

```
In [51]: #merge[['Date', 'Time']] = merge.Date.str.extract(r'\s', expand=True)
```

```
In [52]: X=merge[ 'Date' ].str.split(expand=True)
```

```
In [53]: merge[ 'Time' ]=X[1]
```

```
In [54]: merge[ 'Date' ]=X[0]
```

```
In [55]: merge[ 'Time' ][:4]
```

```
Out[55]: 0    08:45:00
          1    12:00:00
          2    10:30:00
          3    12:04:00
          Name: Time, dtype: object
```

```
In [56]: merge[ 'Date' ][:4]
```

```
Out[56]: 0    11/24/2018
          1    08/27/2018
          2    09/27/2018
          3    10/12/2018
          Name: Date, dtype: object
```

```
In [57]: merge[ 'Month' ]=merge[ 'Date' ].str.slice(0,2)
          merge[ 'Month' ][:3]
```

```
Out[57]: 0    11
          1    08
          2    09
          Name: Month, dtype: object
```

```
In [58]: merge[ 'Day' ]=merge[ 'Date' ].str.slice(3,5)
          merge[ 'Day' ][:2]
```

```
Out[58]: 0    24
          1    27
          Name: Day, dtype: object
```

In [59]:

merge.head()

Out[59]:

|   | District | index | Case Number | Date       | Block                    | IUCR | Primary Type        | Description               | Loc Descri       |
|---|----------|-------|-------------|------------|--------------------------|------|---------------------|---------------------------|------------------|
| 0 | 4        | 2     | JB528220    | 11/24/2018 | 095XX S STONY ISLAND AVE | 0261 | CRIM SEXUAL ASSAULT | AGGRAVATED: HANDGUN       | A                |
| 1 | 22       | 3     | JB495848    | 08/27/2018 | 092XX S UNION AVE        | 1120 | DECEPTIVE PRACTICE  | FORGERY                   | RESIDI           |
| 2 | 7        | 5     | JB505888    | 09/27/2018 | 068XX S PERRY AVE        | 1020 | ARSON               | BY FIRE                   | VEH COMMEF       |
| 3 | 24       | 6     | JB472665    | 10/12/2018 | 061XX N BROADWAY         | 0890 | THEFT               | FROM BUILDING             | COMMEF / BUSI OI |
| 4 | 8        | 9     | JB241744    | 04/27/2018 | 070XX S CALIFORNIA AVE   | 1562 | SEX OFFENSE         | AGG CRIMINAL SEXUAL ABUSE | APARTI           |

5 rows × 28 columns

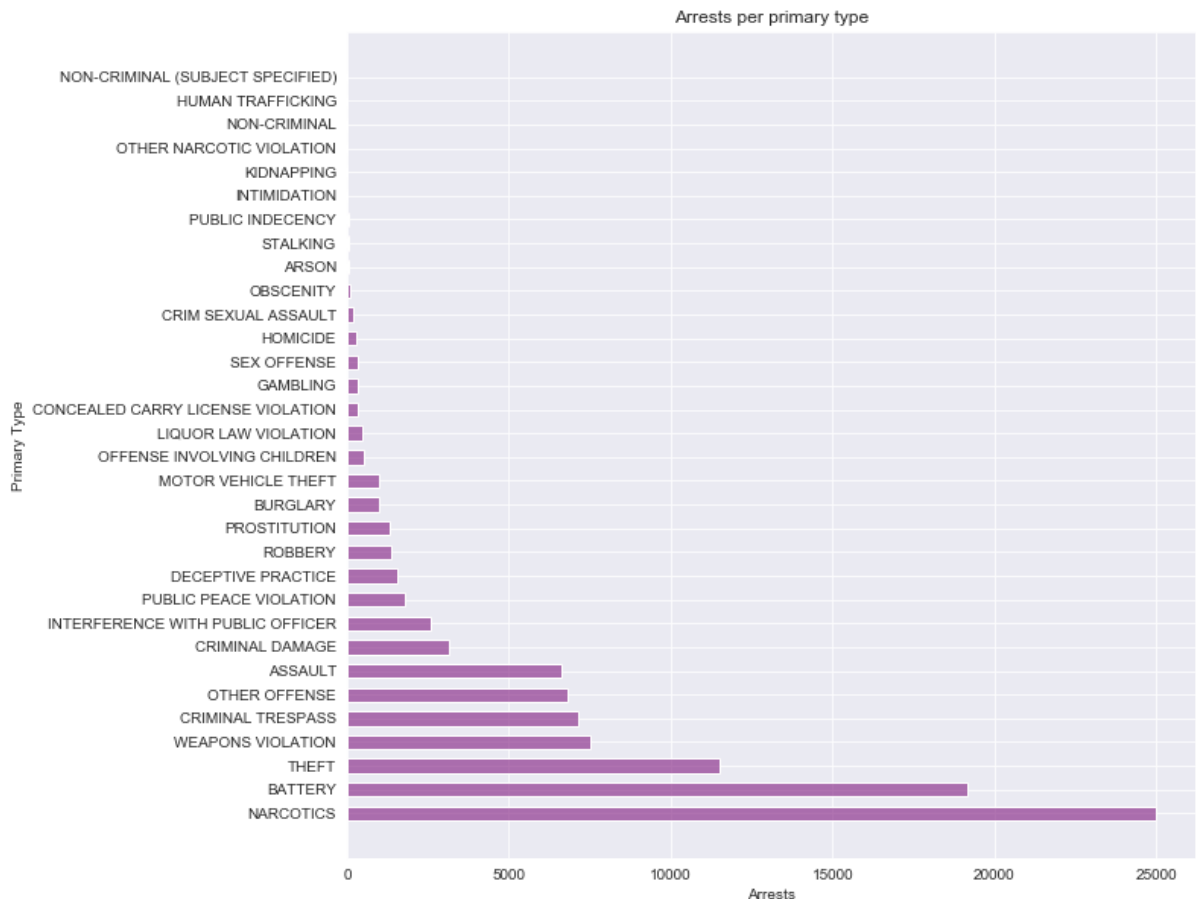
```
In [60]: arrest_true=merge[merge[ 'Arrest' ]==1][ 'Primary Type' ].value_counts().to_
frame()
arrest_true
```

Out[60]:

|  | Primary Type                          |
|--|---------------------------------------|
|  | NARCOTICS 24969                       |
|  | BATTERY 19168                         |
|  | THEFT 11494                           |
|  | WEAPONS VIOLATION 7502                |
|  | CRIMINAL TRESPASS 7153                |
|  | OTHER OFFENSE 6794                    |
|  | ASSAULT 6600                          |
|  | CRIMINAL DAMAGE 3134                  |
|  | INTERFERENCE WITH PUBLIC OFFICER 2566 |
|  | PUBLIC PEACE VIOLATION 1794           |
|  | DECEPTIVE PRACTICE 1554               |
|  | ROBBERY 1343                          |
|  | PROSTITUTION 1293                     |
|  | BURGLARY 984                          |
|  | MOTOR VEHICLE THEFT 969               |
|  | OFFENSE INVOLVING CHILDREN 525        |
|  | LIQUOR LAW VIOLATION 475              |
|  | CONCEALED CARRY LICENSE VIOLATION 338 |
|  | GAMBLING 338                          |
|  | SEX OFFENSE 332                       |
|  | HOMICIDE 279                          |
|  | CRIM SEXUAL ASSAULT 206               |
|  | OBSCENITY 109                         |
|  | ARSON 59                              |
|  | STALKING 46                           |
|  | PUBLIC INDECENCY 23                   |
|  | INTIMIDATION 19                       |
|  | KIDNAPPING 17                         |
|  | OTHER NARCOTIC VIOLATION 2            |
|  | NON-CRIMINAL 2                        |
|  | HUMAN TRAFFICKING 1                   |
|  | NON-CRIMINAL (SUBJECT SPECIFIED) 1    |

```
In [61]: primary_type=merge[merge[ 'Arrest' ]==1][ 'Primary Type' ].value_counts().index.tolist()
arrest_true=merge[merge[ 'Arrest' ]==1][ 'Primary Type' ].value_counts().tolist()
```

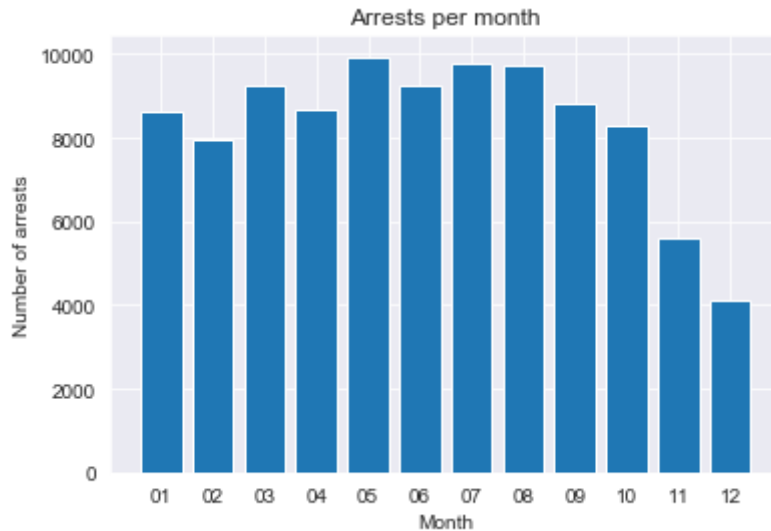
```
In [62]: plt.figure(figsize=(10,10))
plt.barh(primary_type,arrest_true,height=0.6,color = (0.5,0.1,0.5,0.6))
plt.title('Arrests per primary type')
plt.xlabel('Arrests')
plt.ylabel('Primary Type')
plt.yticks(primary_type,fontsize=10)
plt.show()
```



```
In [63]: month=merge[merge[ 'Arrest' ]==1][ 'Month' ].value_counts().sort_index().index.tolist()
arrest_true_m=merge[merge[ 'Arrest' ]==1][ 'Month' ].value_counts().sort_index().tolist()
```

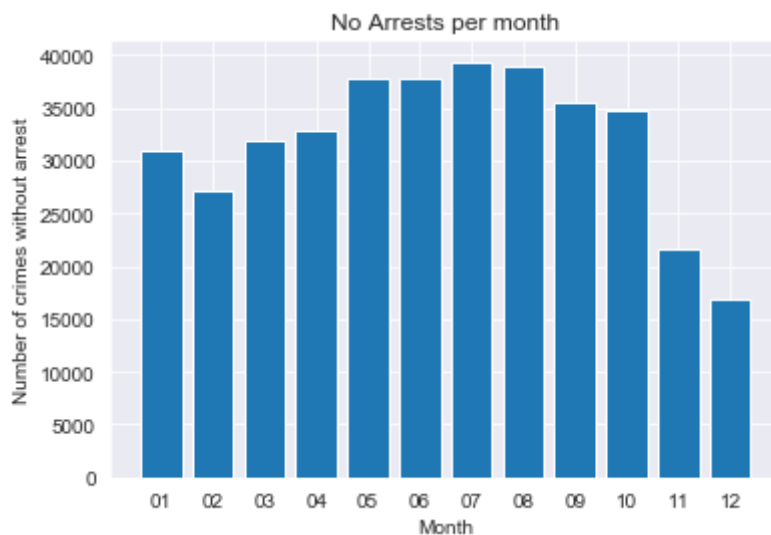


```
In [64]: plt.bar(month,arrest_true_m)
plt.title('Arrests per month')
plt.xlabel('Month')
plt.ylabel('Number of arrests')
plt.show()
```



```
In [65]: noccrime=merge[merge['Arrest']==0]['Month'].value_counts().sort_index().i
ndex.tolist()
arrest_true_m=merge[merge['Arrest']==0]['Month'].value_counts().sort_ind
ex().tolist()
```

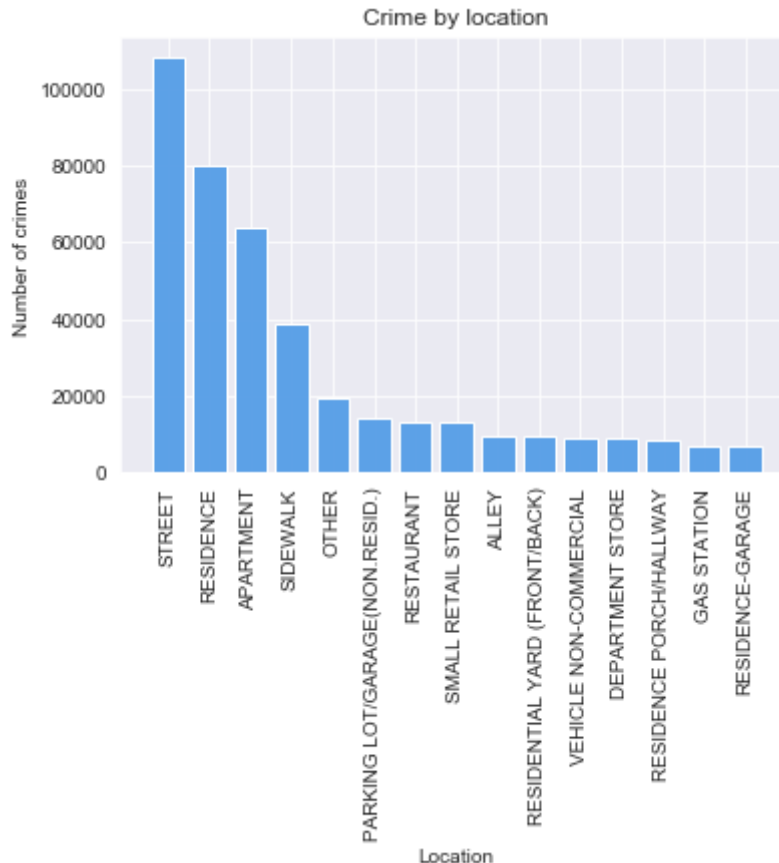
```
In [66]: plt.bar(nocrime,arrest_true_m)
plt.title('No Arrests per month')
plt.xlabel('Month')
plt.ylabel('Number of crimes without arrest')
plt.show()
```



```
In [67]: arrest_true_m
crimes=merge['Location Description'].value_counts().tolist()
location=merge['Location Description'].value_counts().index.tolist()
crimes[0:15]
```

```
Out[67]: [108171,
80271,
63623,
38880,
19643,
14179,
13276,
13215,
9447,
9273,
8977,
8732,
8575,
6939,
6731]
```

```
In [68]: plt.bar(location[0:15],crimes[0:15],color = (0.3,0.6,0.9,0.9))
plt.title('Crime by location')
plt.xlabel('Location')
plt.ylabel('Number of crimes')
plt.xticks(location[0:15],rotation=90)
plt.show()
```

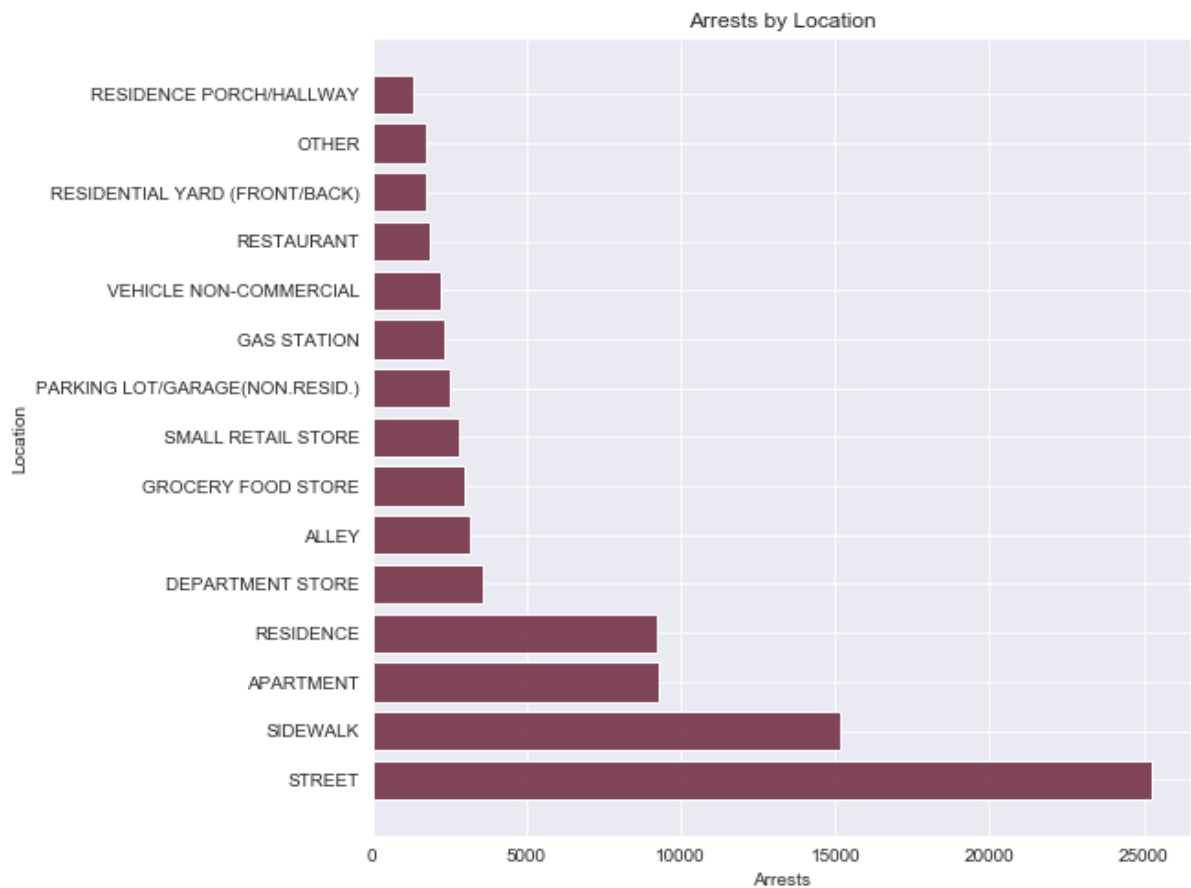


```
In [69]: location=merge[merge['Arrest']==1]['Location Description'].value_counts()  
         ().index.tolist()  
         arrest_loc=merge[merge['Arrest']==1]['Location Description'].value_counts()  
         ().tolist()
```

```
In [70]: location[0:15]
```

```
Out[70]: ['STREET',  
          'SIDEWALK',  
          'APARTMENT',  
          'RESIDENCE',  
          'DEPARTMENT STORE',  
          'ALLEY',  
          'GROCERY FOOD STORE',  
          'SMALL RETAIL STORE',  
          'PARKING LOT/GARAGE(NON.RESID.)',  
          'GAS STATION',  
          'VEHICLE NON-COMMERCIAL',  
          'RESTAURANT',  
          'RESIDENTIAL YARD (FRONT/BACK)',  
          'OTHER',  
          'RESIDENCE PORCH/HALLWAY']
```

```
In [71]: plt.figure(figsize=(8,8))
plt.barh(location[0:15],arrest_loc[0:15],height=0.8,color = (0.4,0.1,0.2
,0.8))
plt.title('Arrests by Location')
plt.xlabel('Arrests')
plt.ylabel('Location')
plt.yticks(location[0:15],fontsize=10)
plt.show()
```



```
In [72]: seasons = {'12': 'Winter', '01': 'Winter', '02' : 'Winter',
                    '03' : 'Spring', '04' : 'Spring', '05' : 'Spring',
                    '06' : 'Summer', '07' : 'Summer', '08' : 'Summer',
                    '09': 'Autumn', '10' : 'Autumn', '11' : 'Autumn'}
merge['Season'] = merge['Month'].apply(lambda x : seasons[x])
merge.head(6)
```

Out[72]:

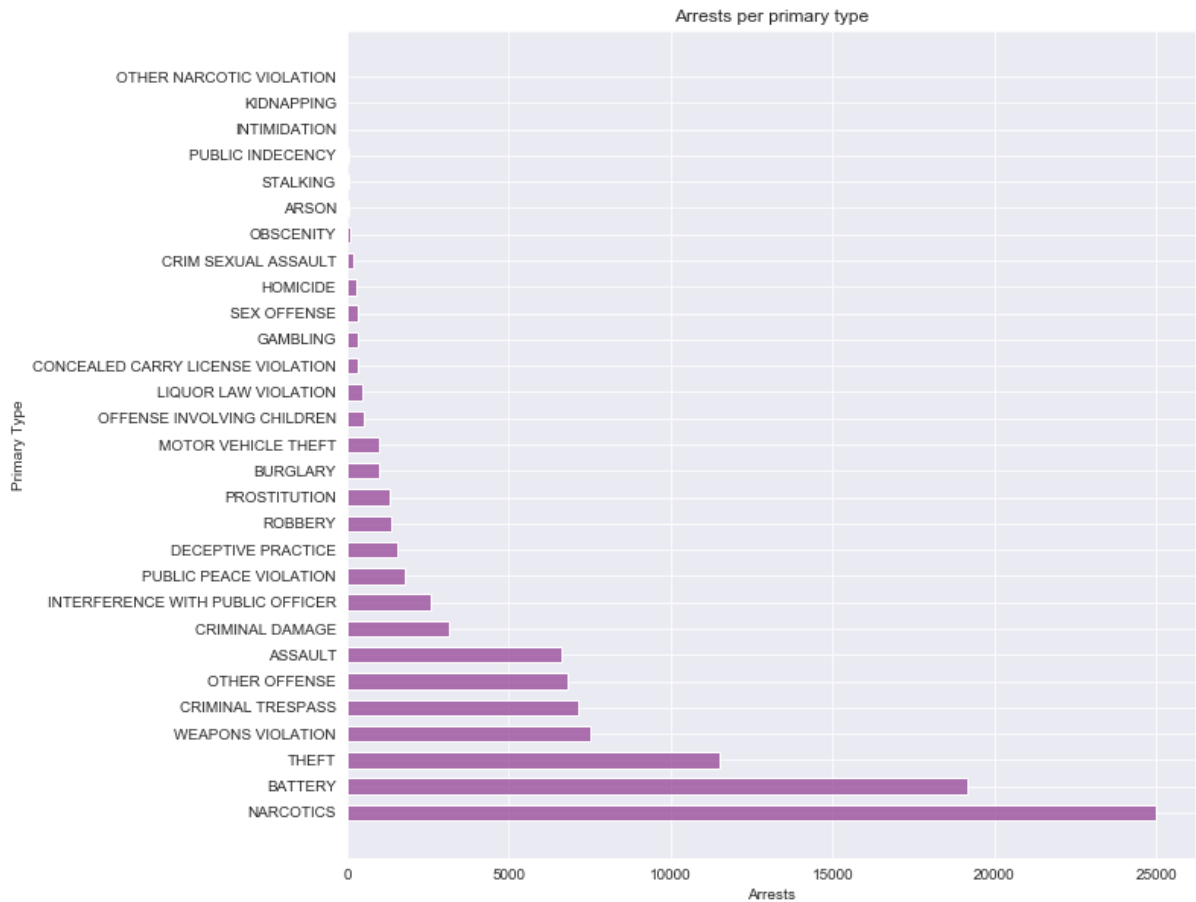
|   | District | index | Case<br>Number | Date       | Block                          | IUCR | Primary<br>Type           | Description                        | Loc<br>Descri          |
|---|----------|-------|----------------|------------|--------------------------------|------|---------------------------|------------------------------------|------------------------|
| 0 | 4        | 2     | JB528220       | 11/24/2018 | 095XX S<br>STONY<br>ISLAND AVE | 0261 | CRIM<br>SEXUAL<br>ASSAULT | AGGRAVATED:<br>HANDGUN             | A                      |
| 1 | 22       | 3     | JB495848       | 08/27/2018 | 092XX S<br>UNION AVE           | 1120 | DECEPTIVE<br>PRACTICE     | FORGERY                            | RESIDI                 |
| 2 | 7        | 5     | JB505888       | 09/27/2018 | 068XX S<br>PERRY AVE           | 1020 | ARSON                     | BY FIRE                            | VEH<br>COMMEF          |
| 3 | 24       | 6     | JB472665       | 10/12/2018 | 061XX N<br>BROADWAY            | 0890 | THEFT                     | FROM<br>BUILDING                   | COMMEF<br>/ BUSI<br>OI |
| 4 | 8        | 9     | JB241744       | 04/27/2018 | 070XX S<br>CALIFORNIA<br>AVE   | 1562 | SEX<br>OFFENSE            | AGG<br>CRIMINAL<br>SEXUAL<br>ABUSE | APARTI                 |
| 5 | 2        | 13    | JB287512       | 01/06/2018 | 053XX S<br>LAKE PARK<br>AVE    | 0460 | BATTERY                   | SIMPLE                             | ST                     |

6 rows × 29 columns

```
In [73]: merge=merge[merge['Primary Type'] != 'NON-CRIMINAL']
merge=merge[merge['Primary Type'] != 'NON-CRIMINAL (SUBJECT SPECIFIED)']
merge=merge[merge['Primary Type'] != 'HUMAN TRAFFICKING']
```

```
In [74]: primary_type=merge[merge['Arrest']==1]['Primary Type'].value_counts().in
dex.tolist()
arrest_true=merge[merge['Arrest']==1]['Primary Type'].value_counts().tol
ist()
```

```
In [75]: plt.figure(figsize=(10,10))
plt.barh(primary_type,arrest_true,height=0.6,color = (0.5,0.1,0.5,0.6))
plt.title('Arrests per primary type')
plt.xlabel('Arrests')
plt.ylabel('Primary Type')
plt.yticks(primary_type,fontsize=10)
plt.show()
```



```
In [76]: crimeGroups = {'NARCOTICS' : '3', 'BATTERY' : '1', 'THEFT': '2', 'WEAPONS  
VIOLATION': '5', 'CRIMINAL TRESPASS' : '2',  
'OTHER OFFENSE' : '4', 'ASSAULT' : '1', 'CRIMINAL DAMAGE' : '2',  
'INTERFERENCE WITH PUBLIC OFFICER' : '1',  
'PUBLIC PEACE VIOLATION' : '1',  
'DECEPTIVE PRACTICE' : '3',  
'ROBBERY' : '2',  
'PROSTITUTION' : '3',  
'BURGLARY' : '2',  
'MOTOR VEHICLE THEFT' : '3',  
'OFFENSE INVOLVING CHILDREN' : '3',  
'LIQUOR LAW VIOLATION' : '3',  
'CONCEALED CARRY LICENSE VIOLATION' : '5',  
'GAMBLING' : '3',  
'SEX OFFENSE' : '1',  
'HOMICIDE' : '1',  
'CRIM SEXUAL ASSAULT' : '1',  
'OBSCENITY' : '1',  
'ARSON' : '2',  
'STALKING' : '1',  
'PUBLIC INDECENCY' : '3',  
'INTIMIDATION' : '1',  
'KIDNAPPING' : '1',  
'OTHER NARCOTIC VIOLATION' : '3'}  
  
merge['crimeGroups'] = merge['Primary Type'].apply(lambda x : crimeGroup  
s[x])  
merge.head(6)  
# Personal crime = physical or mental to there person = 1  
# Property crimes = Property involved [Robbery, arson] = 2  
# Crime against morality = drugs, prostitutions, gambling = 3  
# Other = 4  
# Firearms = 5
```

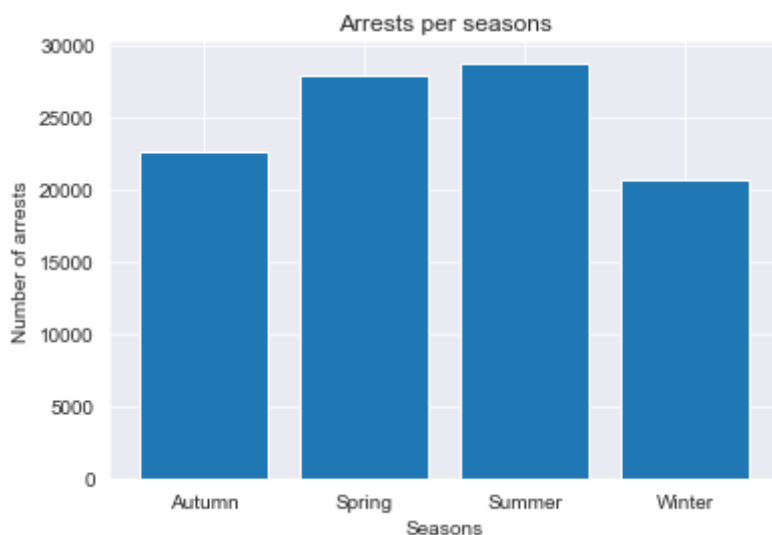
Out[76]:

|   | District | index | Case Number | Date       | Block                    | IUCR | Primary Type        | Description               | Loc Descri       |
|---|----------|-------|-------------|------------|--------------------------|------|---------------------|---------------------------|------------------|
| 0 | 4        | 2     | JB528220    | 11/24/2018 | 095XX S STONY ISLAND AVE | 0261 | CRIM SEXUAL ASSAULT | AGGRAVATED: HANDGUN       | A                |
| 1 | 22       | 3     | JB495848    | 08/27/2018 | 092XX S UNION AVE        | 1120 | DECEPTIVE PRACTICE  | FORGERY                   | RESIDI           |
| 2 | 7        | 5     | JB505888    | 09/27/2018 | 068XX S PERRY AVE        | 1020 | ARSON               | BY FIRE                   | VEH              |
| 3 | 24       | 6     | JB472665    | 10/12/2018 | 061XX N BROADWAY         | 0890 | THEFT               | FROM BUILDING             | COMMEF / BUSI OI |
| 4 | 8        | 9     | JB241744    | 04/27/2018 | 070XX S CALIFORNIA AVE   | 1562 | SEX OFFENSE         | AGG CRIMINAL SEXUAL ABUSE | APARTI           |
| 5 | 2        | 13    | JB287512    | 01/06/2018 | 053XX S LAKE PARK AVE    | 0460 | BATTERY             | SIMPLE                    | ST               |

6 rows × 30 columns

```
In [77]: season=merge[merge[ 'Arrest' ]==1][ 'Season' ].value_counts().sort_index().i
ndex.tolist()
arrest_true_m=merge[merge[ 'Arrest' ]==1][ 'Season' ].value_counts().sort_in
dex().tolist()
```

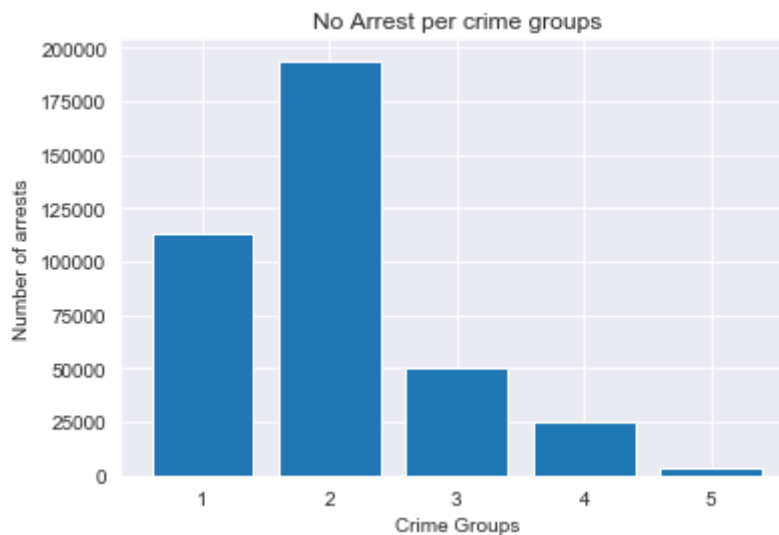
```
In [78]: plt.bar(season,arrest_true_m)
plt.title('Arrests per seasons')
plt.xlabel('Seasons')
plt.ylabel('Number of arrests')
plt.show()
```





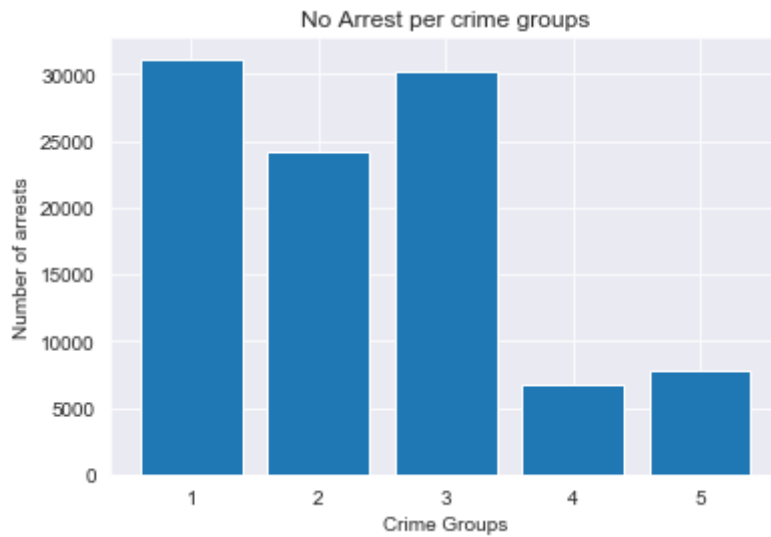
```
In [79]: crimenGroups=merge[merge['Arrest'] == 0]['crimeGroups'].value_counts().  
sort_index().index.tolist()  
arrest_false_m=merge[merge['Arrest'] == 0]['crimeGroups'].value_counts().  
sort_index().tolist()
```

```
In [80]: plt.bar(crimenGroups,arrest_false_m)  
plt.title('No Arrest per crime groups')  
plt.xlabel('Crime Groups')  
plt.ylabel('Number of arrests')  
plt.show()
```



```
In [81]: crimenGroups=merge[merge['Arrest'] == 1]['crimeGroups'].value_counts().  
sort_index().index.tolist()  
arrest_true_m=merge[merge['Arrest'] == 1]['crimeGroups'].value_counts().  
sort_index().tolist()
```

```
In [82]: plt.bar(crimenGroups,arrest_true_m)
plt.title('No Arrest per crime groups')
plt.xlabel('Crime Groups')
plt.ylabel('Number of arrests')
plt.show()
```



```
In [83]: merge.to_csv('Merge.csv',index=None)
```

```
In [ ]:
```

```
In [ ]:
```